

Individual differences in VOT realizations of English plosives by Palestinian pupils

Nour Tamim, Silke Hamann, Paul Boersma

University of Amsterdam
 {n.j.s.tamim, silke.hamann, paul.boersma}@uva.nl

ABSTRACT

Palestinian Arabic (PA) learners of English meet with two types of challenges when learning English plosives in initial position. The first challenge is that English initial /p, t, k/ are aspirated while PA /t, k/ are not, and that English /b, d, g/ have short-lag VOT while PA /b, d/ are fully voiced. The second challenge is that PA lacks /p/ and (to some extent) /g/. This paper focuses on individual differences between learners in how they cope with the two types of challenges. As could be expected, some learners have fully voiced /b, d, g/ in their English, and/or unaspirated /p, t, k/. As could also be expected, some learners pronounce /p/ and /b/ in the same way. We found no evidence of an implicational hierarchy between these challenges, e.g. whether learners who conflate /p/ and /b/ also have fully voiced /b, d, g/ in their English.

Keywords: VOT, second language acquisition, variation, Palestinian Arabic, English.

1. INTRODUCTION

There are several differences between English and Palestinian Arabic (PA) when it comes to the structure and implementation of the plosive voicing contrasts in these two languages.

As far as structure is concerned, English contrasts voiced and voiceless plosives at three places of articulation, namely bilabial (/p, b/), alveolar (/t, d/), and velar (/k, g/). Palestinian Arabic (PA), on the other hand, contrasts denti-alveolar stops (/t, d/) and their pharyngeal counterparts (/t^ʕ, d^ʕ/), but at the bilabial and velar places of articulation, there are gaps: while PA has /b/ and /k/, it lacks /p/ and /g/ (though these do appear in loanwords, and /g/ in some of the dialects). Both gaps are quite common cross-linguistically (see [1] for an overview).

English and PA also lie in different corners of the typology of the phonetic realization of voicing contrasts established by Lisker & Abramson [2], as measured on the voice onset time (VOT) continuum. In initial position, PA employs voicing lead for /b, d/ and short lag for /t, k/ [3], while English employs short lag (with occasional prevoicing) for /b, d, g/ and long lag for /p, t, k/ [2]. In other words, PA contrasts

prevoiced plosives with unaspirated plosives, whereas English contrasts unaspirated plosives with aspirated plosives.

Differences in structure and implementation of voicing may provide obstacles in second-language learning. Incomplete acquisition of an L2 voicing contrast was observed in learners of English whose L1 utilizes prevoicing to distinguish voiced plosives from voiceless ones. Such learners typically employ their L1 implementation of voicing (namely, voicing lead) when realizing English voiced plosives; this has been found for Saudi Arabic speakers [4], Dutch speakers [5], and Italian speakers [6].

The present study looks at inter-speaker variation in the production of initial English plosive voicing by Palestinian Arabic teenagers who have learned English from a native English teacher at school: how do they handle the difference between the voicing structures of PA and English, how do they handle the difference between the phonetic realization of voicing in PA and English, and how is their handling of structure related to their handling of phonetic realization?

2. VOICING MEASUREMENT

It would be ideal if the voicing continuum were simply one-dimensional, i.e. if the degree of voicing could be determined by measuring a single VOT value. Davidson [7] pointed out that this will not work for plosives that have voicing during their closure period: cases of a true negative VOT, i.e. a closure that may start out as voiceless, then switches to voiced, and stays voiced during the remainder of the closure, turn out to be rare in English. Instead, for instance, *voicing bleed* from the preceding vowel (or other sonorant) turns out to be common. Davidson therefore proposed a detailed classification scheme, which takes into consideration partial voicing in the constriction interval of plosives.

We loosely follow Davidson's classification procedure, with the goal of establishing an ordering of phonetic voicedness (disregarding aspiration), as follows.

For each token of an initial plosive (spoken by a PA learner of English) we label the beginning and end of the closure phase manually in Praat [8]. The start of the closure is determined by the offset of the

preceding vowel, and its end by the start of the release of the plosive. We then divide the closure into three equal parts. For each part we determine the percentage of voicing with the help of Praat’s Voice Report (*fraction of locally unvoiced frames*), which uses Boersma’s [9] pitch detection algorithm, with the silence threshold set to 0.02 and the voicing threshold set to 0.6 (to avoid that occasional background noise is interpreted as voicing).

We consider a closure part “voiced” if the Voice Report reports 50 or more percent voicing, and “voiceless” otherwise. This yields the eight possible voicing patterns listed in Table 1. To establish an ordering to work with, we classify the eight patterns into three degrees of closure voicing (similarly though not identically to [7]): we classify a closure as *voiced* if more than 50% voicing is detected in all three parts or in the first two parts (the latter are cases where vocal fold vibration ceases in the last third, usually due to lack of transglottal pressure differences [10]); we classify a closure as *voiceless* if less than 50% is detected in all three parts or in the last two parts (in the latter case, the voicing in the first part is considered a short bleeding effect from the preceding vowel); we classify the remaining four cases as *partially voiced*. For a full overview, see Table 1.

first	second	third	class label
V	V	V	voiced
V	V	–	
V	–	V	partially voiced
–	V	V	
–	V	–	
–	–	V	
V	–	–	voiceless
–	–	–	

Table 1: Classification based on percentage voicing in the first, second, and third part of the closure phase; “–” indicates less than 50% voicing in the part, “V” 50% or more voicing.

3. EXPERIMENT

3.1. Participants

Participants were 20 pupils of the 10th grade (16 years

of age) at a school in Nablus where English is taught by native English speakers. They had had 10 years of English classes at school.

(The present results are part of a larger study that looks at the influence of input by native versus non-native English teachers on the English production of Palestinian students [11].)

3.2. Material and procedure

18 English words that students are well-familiar with were used as stimuli in a picture naming task. All words had an initial plosive, and were followed by one of three vowels, see the overview in Table 2.

	/b/	/d/	/g/	/p/	/t/	/k/
/a:/	basket	dark	garden	palm	tart	car
/u:/	boots	do	goose	pool	two	cool
/i:/	bee	D	geese	peas	tea	key

Table 2: English test words with initial plosives in three different following vowel contexts.

Speakers were presented with the picture of a word three times, they thus produced three repetitions of each word and a total of 54 stimuli. The stimuli and 26 filler words were presented in randomized order. Each word had to be produced in the carrier sentence *This is a _*, causing an intervocalic context for the plosive.

Instructions were given in English. The recordings took place in a quiet room with an Edirol R-1 solid-state recorder, an Intricon HT-747 super-cardioid electret condenser head-mounted microphone, and a sample frequency of 44.1 kHz.

4. RESULTS

Of the total of 1080 tokens (18 test words × 20 speakers × 3 repetitions), 16 tokens had to be excluded, mainly because of background noise. The distribution of closure voicing classes for the remaining 1064 tokens is given in Table 3. We can see that the class “partially voiced” was rare for all plosives.

Figures 1–3 provide a visual overview of the speaker-specific use of the three voicing classes.

classes	/b/		/d, g/		/t, k/		/p/	
voiced	147	83%	253	71%	10	3%	37	21%
partially voiced	8	4%	13	4%	24	7%	9	5%
voiceless	23	13%	90	25%	320	90%	130	74%
total	178		356		354		176	

Table 3: Occurrence of different classes of voicing realization during closure phases.

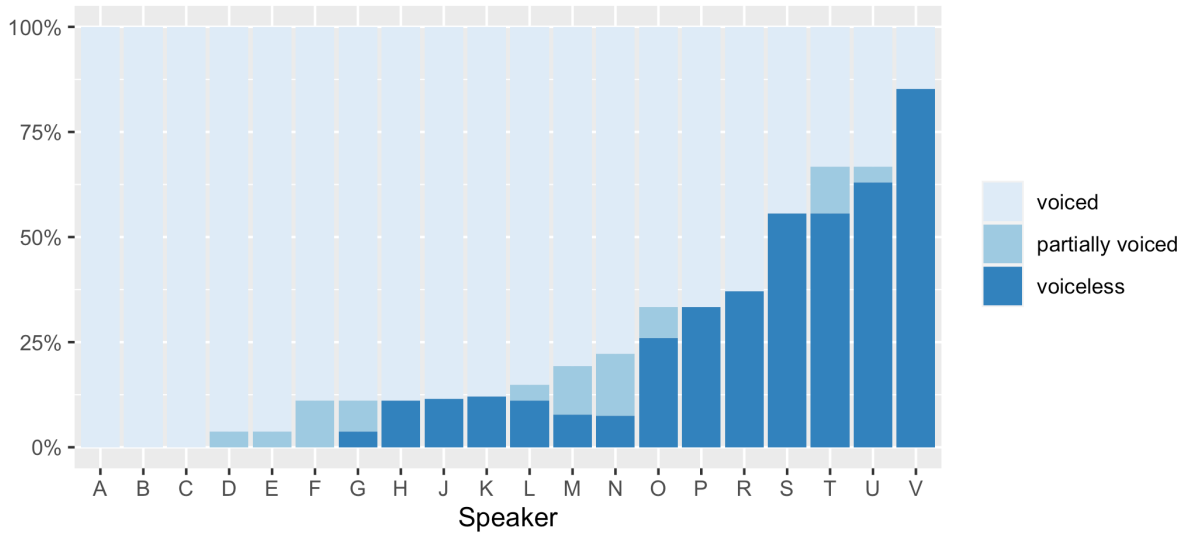


Figure 1: Speaker-specific realizations of the closure phase in English voiced plosives.

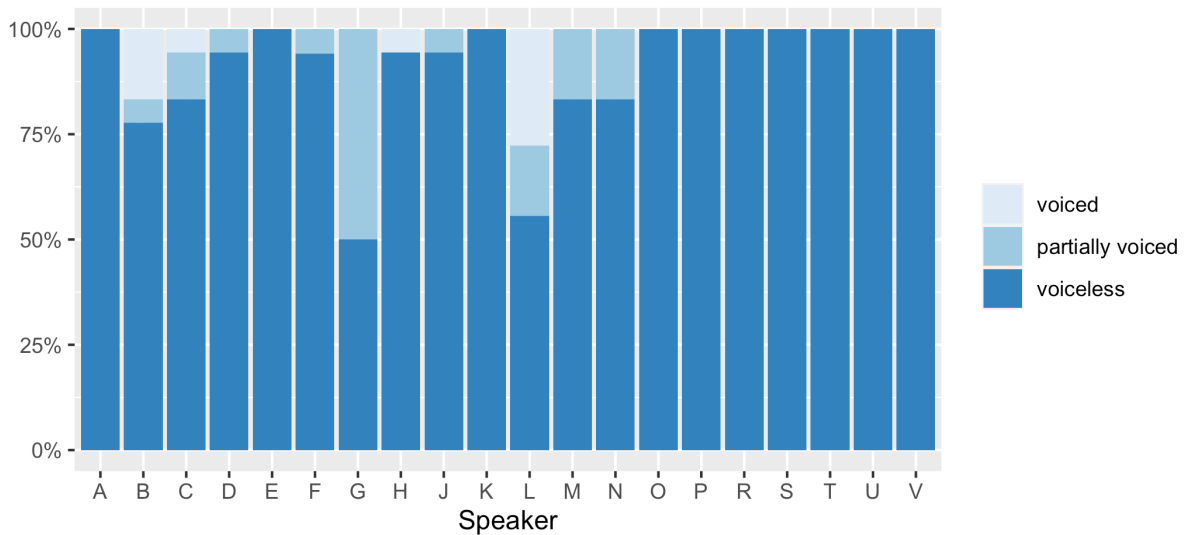


Figure 2: Speaker-specific realizations of the closure phase in English /t/ and /k/.

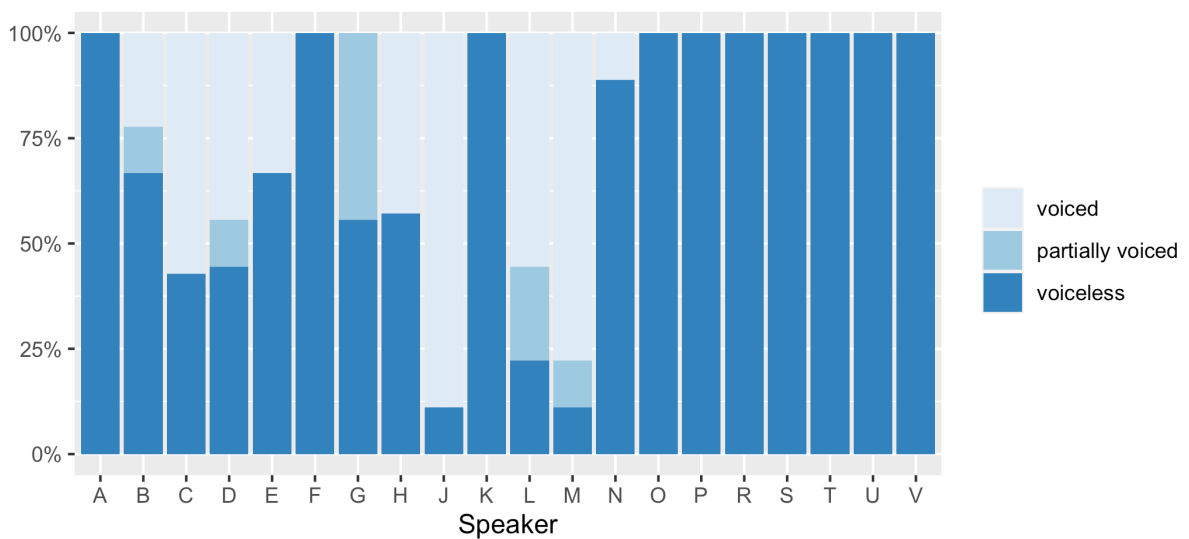


Figure 3: Speaker-specific realizations of the closure phase in English /p/.

For **statistical analysis**, we constructed a mixed-effects model in *R* [12] using the package *lme4* [13], with the *closure voicing* class as an ordered dependent variable that we assumed linear: we coded *voiceless* as -0.5 , *partially voiced* as 0 , and *voiced* as $+0.5$. The model contained two nominal within-speaker fixed effects: the target (English) phonological features [*place*] (of articulation) and [*voice*]. For [*place*], the research question about the lack of a phonetic voicing distinction between /b/ and /p/ involves a ternary contrast between *labial* (coded as $+2/3$) on the one hand, and *coronal* and *dorsal* (both coded as $-1/3$) on the other hand, supplemented with a binary contrast between *coronal* (coded as $+0.5$) and *dorsal* (coded as -0.5). For [*voice*], there is only a binary contrast between [$-voi$] (coded as -0.5) and [$+voi$] (coded as $+0.5$). All contrasts are thus orthogonal and sum to zero, and they differ by 1.0 so that the model's estimates will be interpretable as effect sizes. Computing the model involved fitting the following parameters: four fixed effects (intercept, [*place*], [*voice*], and the interaction of [*place*] and [*voice*]); per speaker a random intercept as well as random slopes for [*place*], [*voice*], and [*place*] \times [*voice*]; and per word a random intercept.

As expected (and included here solely as a sanity check), the model reported an estimated **[voice] effect** of $+0.65$, i.e. the difference between [$+voi$] and [$-voi$] consonants amounts to 65 percent of the closure voicing scale. Since $t = 10.155$, we conclude that (for 16-year-old PA learners of English with native-English teachers) the [$+voi$] consonants /b, d, g/ tend to have more closure voicing (as a group) than the [$-voi$] consonants /p, t, k/ (Satterthwaite's approximation yields $df = 27.25$, $p = 9.2 \cdot 10^{-11}$, and 95% CI = $+0.52 \dots +0.78$). This means that the visual distinction between Figure 1 on the one hand and Figures 2 and 3 on the other hand reflects a true population effect, and that this effect is large (at least 52 percent of the closure voicing scale).

As also expected, /p, b/ taken together have more closure voicing than /t, d, k, g/ taken together. The interval estimate for the **ternary [place] effect** runs from $+0.03$ to $+0.27$, i.e. we can conclude that labials on average have between 3 and 27 percent points more closure voicing than non-labials (point estimate $+0.15$, $t = 2.660$, $df = 20.52$, $p = 0.015$). This means that the visual distinction between Figures 2 and 3 (which according to Fig. 1 cannot exist to this extent for the [$+voi$] plosives) reflects a true population effect, and this effect can be small or medium (between 3% and 27% of the closure voicing scale).

If this labiality effect reflects the existence of a smaller distinction between /b/ and /p/ than between

/d/ and /t/ or between /g/ and /k/, we can hope for a negative interaction effect between the ternary [*place*] contrast and [*voice*]. The model did not, however, reliably detect this (CI = $-0.36 \dots +0.23$).

Differences between coronals and dorsals were detected neither for the binary [*place*] effect (CI = $-0.12 \dots +0.13$) nor for its interaction with the [*voice*] effect (CI = $-0.21 \dots +0.27$).

Individual differences. When visually comparing all Figures 1–3, it seems that speakers J and M, and perhaps L, implement /p/ as equally voiced as the [$+voi$] plosives. The only way in which our mixed-effects model can hint at the existence of such differences of strategies between speakers lies in the standard deviations of the random effects: the between-speaker standard deviation is highest (48 percent of the closure voicing scale) for the interaction between ternary [*place*] and [*voice*], and lower (between 8 and 21 percent) for the other five by-speaker random effects.

5. DISCUSSION

Besides an obvious expected result ([$+voi$] plosives have more closure voicing than [$-voi$] plosives), we have found several effects that can be ascribed to differences between Palestinian Arabic and English.

The structural challenge of PA lacking /p/ could lead to /p/ merging into /b/ for some learners. This phenomenon can lie behind the detected ternary [*place*] effect ($p = 0.015$), as well as (tentatively) behind the large between-speaker standard deviation for the interaction between [*voice*] and ternary [*place*]. Future modelling with a latent predictor (L2 phonological merger of /p/ with /b/) may shed more light on this issue. It is already clear from the 10 speakers who totally devoice /p/ in Figure 3, several of whom must have full closure voicing for all [$+voi$] plosives according to Figure 1, that there are speakers who do master the English /p~/b/ opposition.

The only phonetic implementation challenge we studied was the English devoicing of initial [$+voi$] plosives. According to Figure 1, a few PA learners master this, but most don't, even after 10 years of native input.

We found no evidence for a hierarchical or conditional or implicational relationship between these two challenges. Both challenges might be remedied by specific minimal-pair training [14, 15].

We conclude that Palestinian Arabic learners of English vary in the extent to which they carry over their native sound structure to their L2 (some merge /p/ with /b/, some don't), as well in the extent to which they carry over their native phonetic implementation.

6. REFERENCES

- [1] Seinhorst, K. 2021. *The Learnability and Complexity of Phonological Patterns: Simulations, Experiments, Typology*. Doctoral thesis, University of Amsterdam.
- [2] Lisker, L., Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *WORD*, 20(3), 384–422.
- [3] Tamim, N., Hamann, S. 2021. Voicing contrasts in the singleton stops of Palestinian Arabic: Production and perception. *Proc. Interspeech 2021* Brno, 401–405. doi: 10.21437/Interspeech.2021-1079
- [4] Flege, J.E. & Port, R. 1981. Cross-language phonetic interference: Arabic to English. *Language and Speech* 24(2), 125–146.
- [5] Simon, E. 2009. Acquiring a new second language contrast: An analysis of the English laryngeal system of native speakers of Dutch. *Second Language Research* 25(3), 377–408.
- [6] MacKay, I.R.A., Flege, J.E., Piske, T., Schirru, C. 2001. Category restructuring during second-language speech acquisition. *J. Acoust. Soc. Am.* 110(1), 516–528.
- [7] Davidson, L. 2016. Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics* 54, 35–50.
- [8] Boersma, P., Weenink, D. 1992–2022. *Praat: Doing Phonetics by Computer*. <https://www.praat.org>.
- [9] Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences Amsterdam* 17, 97–110.
- [10] Klatt, D.H. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* 18(4), 686–706.
- [11] Tamim, N., Hamann, S. 2023. The influence of native speaker input on the production of the English voicing contrast in stops by Palestinian Arabic pupils. Manuscript, University of Amsterdam.
- [12] R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org>.
- [13] Bates, D., Mächler, M., Bolker, B, Walker, S. 2015. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* 67(1), 1–48.
- [14] Barlow, J.A., Gierut, J.A. 2002. Minimal pair approaches to phonological remediation. *Seminars in Speech and Language* 23(1), 57–68.
- [15] Hamzah, M. H., Bawodood, A. M. A. 2019. Teaching English sounds via minimal pairs: The case of Yemeni EFL learners. *Veda's Journal of English Language and Literature* 6(3), 97–102.